DISCRIMINANT ANALYSIS FOR MULTIVARIATE NON-GAUSSIAN LOCALLY STATIONARY PROCESSES

JUNICHI HIRUKAWA∗

Received February 5, 2004; revised March 3, 2004

Abstract. The extension of classical discriminant analysis techniques in multivariate analysis to time series data is a problem of practical interest. Discrimination between different classes of multivariate locally stationary processes, which constitute a class of non-stationary processes, can be characterized by differing covariance or time varying spectral structures. For discrimination between the multivariate non-Gaussian locally stationary processes, Kullback-Leibler discrimination information measure has been developed. In this paper, asymptotic error rates and limiting distributions are given for a generalized time varying spectral disparity measure that includes foregoing criteria as special case. It is well known that the log-likelihood ratio based on observed stretch gives optimal classification. It is shown that the discriminant criterion based on such generalized disparity measure is Gaussian optimal. A non-Gaussian optimal discriminant criterion is also proposed in view of the LAN theorem.

1. Introduction. In multivariate analysis, many methods of discriminant analysis have been investigated in detail (e.g. Anderson, 1984). The extension of classical discriminant analysis techniques in multivariate analysis to time series data is a problem of practical interest. Shumway (1982) gave an extensive review of various discriminant problems in time series. Zhang and Taniguchi (1994) discussed the parametric discriminant problems for non-Gaussian vector linear processes, and showed that discriminant criterion based on an integral functional of periodogram has some good properties, for example, asymptotic normality and non-Gaussian robustness, etc. Zhang and Taniguchi (1995) have shown robustness of Chernoff information measure to peak contamination in spectra of the process concerned. For discrimination between non-Gaussian multivariate time series, Kakizawa et al. (1998) have introduced a disparity measure, which includes the Kullback-Leibler discrimination information and the Chernoff information measure, and gave applications to the problems of classifying earthquakes and mining explosions.

Although the analysis for stationary time series is well established, stationary time series models are not plausible to describe the real world. Dahlhaus (1996a, 1996b, 1996c, 1997) has introduced an important class of non-stationary processes, called locally stationary processes, and established the asymptotic theory of statistical inference. Discrimination between different classes of multivariate time series that can be characterized by differing covariance or time varying structures is important in applications of occurring in the analysis of seismic records and biometrics data. Sakiyama and Taniguchi (2004) investigated the problems of classifying a multivariate non-Gaussian locally stationary process \( \{X_{t,T}\} \) into one of two categories described by two hypotheses \( \Pi_i : f_i(u, \lambda), i = 1, 2 \), where \( f_i(u, \lambda) \) are time varying spectral density matrices. They used an approximation of the Gaussian Kullback-Leibler information measure as a classification statistic for this problem and showed that this statistic is consistent. The misclassification probabilities were

2000 Mathematics Subject Classification. 62M10; 62M15; (62H30).

Key words and phrases. Locally stationary vector process; Discriminant analysis; Chernoff; Kullback-Leibler; Non-Gaussian robust; Peak robustness; Non-Gaussian optimal.
also evaluated under contiguous hypotheses. In this paper, we generalize this measure to non-linear time varying spectral measures which include the Kullback-Leibler information and the Cernoff information measure. We also propose a genuine non-Gaussian optimal discrimination criterion based on another approach. The time series data recorded in real phenomena such as seismic record and financial time series, are often non-stationary and non-Gaussian. To investigate the actual performance of our discrimination criterion to such multivariate non-stationary and non-Gaussian time series data will be increasing importance. However, this problem requires another paper. We will make it as a future work.

This paper is organized as follows. In Section 2 we define the multivariate non-Gaussian locally stationary processes, and introduce a nonparametric time varying spectral density estimator, which is due to Dahlhaus (1996a, 1996b, 1997). Section 3 gives a generalized measure of disparity which includes Kullback-Leibler and Chernoff information measure. In Section 4, we derive the limiting distributions and asymptotic error rates of our discriminant statistics. We also discuss conditions for non-Gaussian robustness, and show that our discriminant criterion is Gaussian optimal. Peak robustness of our discrimination criterion is studied, and some numerical examples are given. In Section 5, we propose a genuine non-Gaussian optimal discrimination criterion based on the LAN property. All the proofs of Theorems are given in Section 6.

2. Non parametric spectral estimator of multivariate locally stationary processes. When we deal with nonstationary processes, one of the difficult problems to solve is how to set up an adequate asymptotic theory. To meet this Dahlhaus (1996a, 1996b, 1997) introduced an important class of nonstationary processes and developed the statistical inference. We give the precise definition of multivariate locally stationary processes which is due to Dahlhaus (2000).

Definition 1. A sequence of multivariate stochastic processes $X_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(m)})'$, $(t = 2 - N/2, \ldots, 1, \ldots, T, \ldots, T + N/2; T, N \geq 1)$ is called locally stationary with mean vector 0 and transfer function matrix $\mathbf{A}^\circ$ if there exists a representation

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) \mathbf{A}^\circ_{t,T}(\lambda) d\xi(\lambda),$$

where

(i) $\xi(\lambda) = (\xi_1(\lambda), \ldots, \xi_m(\lambda))'$ is a complex valued stochastic vector process on $[-\pi, \pi]$ with $\xi_a(\lambda) = \xi_a(-\lambda)$ and

$$(2) \quad \text{cum}\{d\xi_{a_1}(\lambda_1), \ldots, d\xi_{a_k}(\lambda_k)\} = \eta(\sum_{j=1}^{k} \lambda_j) \frac{k^{a_1 \ldots a_k}}{(2\pi)^{k-1}} d\lambda_1 \ldots d\lambda_{k-1},$$

for $k \geq 2$, $a_1, \ldots, a_k = 1, \ldots, m$, where $\text{cum}\{\ldots\}$ denotes the cumulant of $k$-th order, and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period $2\pi$ extension of the Dirac delta function.

(ii) There exists a constant $K$ and a $2\pi$-periodic matrix valued function $\mathbf{A} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}^{m \times m}$ with $\mathbf{A}(u, -\lambda) = \mathbf{A}(u, \lambda)$ and

$$(3) \quad \sup_{t,\lambda} \left| \mathbf{A}^\circ_{t,T}(\lambda)_{a,b} - \mathbf{A} \left( \frac{t}{T}, \lambda \right)_{a,b} \right| \leq KT^{-1}$$

for all $a, b = 1, \ldots, m$ and $T \in \mathbb{N}$. $\mathbf{A}(u, \lambda)$ is assumed to be continuous in $u$. 
\[ f(u, \lambda) := A(u, \lambda) \Omega A(u, \lambda)^* \] is called the time varying spectral density matrix of the process, where \( \Omega = (\kappa_{a,b})_{a,b=1,\ldots,m} \) and \( D^* \) denotes the complex conjugate of matrix \( D \). Write

\[ \varepsilon_t := \int_{-\pi}^{\pi} \exp(i\lambda t) d\xi(\lambda), \]

then \( E(\varepsilon_t) = 0 \), \( E(\varepsilon_t \varepsilon_t') = \Omega \) and \( E(\varepsilon_t \varepsilon_s') \) for \( t \neq s \) is a zero matrix. We make the following assumption.

**Assumption 1.** \( X_{t,T} \) has the \( MA(\infty) \) representation

\[ X_{t,T} = \sum_{s=-\infty}^{\infty} a_{t,T}(s) \varepsilon_{t-s}, \]

that is,

\[ A_{t,T}(\lambda) = \sum_{s=-\infty}^{\infty} a_{t,T}(s) \exp(-i\lambda s), \]

where the coefficients fulfill

\[ \sup \sum_{t,s=-\infty}^{\infty} \left| \begin{pmatrix} a_{t,T}(s) - a_s \left( \frac{t}{T} \right) \end{pmatrix} \right|_{cd} = O(T^{-1}) \]

for all \( c, d = 1, \ldots m \), with continuous matrix function \( a_s(u) \). Then, the condition (3) is fulfilled for

\[ A(u, \lambda) = \sum_{s=-\infty}^{\infty} a_s(u) \exp(-i\lambda s). \]

Furthermore we make the following assumption on the transfer function matrix \( A(u, \lambda) \).

**Assumption 2.** (i) The transfer function matrix \( A(u, \lambda) \) is \( 2\pi \)-periodic in \( \lambda \) and the periodic extension is twice differentiable in \( u \) and \( \lambda \) with uniformly bounded continuous derivatives \( \frac{\partial^2}{\partial u^2} A \), \( \frac{\partial^2}{\partial \lambda^2} A \) and \( \frac{\partial^2}{\partial u^2} \frac{\partial}{\partial \lambda} A \).

(ii) All the eigenvalues of \( f(u, \lambda) \) are bounded from below and above by some constants \( \delta_1, \delta_2 > 0 \) uniformly in \( u \) and \( \lambda \).

As an estimator of \( f(u, \lambda) \), we use the nonparametric estimator of kernel type defined by

\[ \hat{f}_T(u, \lambda) = \int_{-\pi}^{\pi} W_T(\lambda - \mu) I_N(u, \mu) d\mu, \]

where \( W_T(\omega) = M \sum_{\nu=-\infty}^{\infty} W(M(\omega + 2\pi\nu)) \) is the weight function and \( M > 0 \) depends on \( T \), and \( I_N(u, \lambda) \) is the data tapered periodogram matrix over the segment \( \{[uT] - N/2 + 1, [uT] + N/2 \} \) defined as

\[ I_N(u, \lambda) = \frac{1}{2\pi H_{2,N}} \left\{ \sum_{s=1}^{N} h \left( \frac{s}{N} \right) X_{[uT]-N/2+s,T} \exp\{i\lambda s\} \right\} \cdot \left\{ \sum_{r=1}^{N} h \left( \frac{r}{N} \right) X_{[uT]-N/2+r,T} \exp\{i\lambda r\} \right\}^*. \]
Here \( h : [0, 1] \rightarrow \mathbb{R} \) is a data taper and \( H_{2, N} = \sum_{s=1}^{N} h(s) \). It should be noted that \( I_{N}(u, \lambda) \) is not a consistent estimator of the spectral density. To make a consistent estimator of \( f(u, \lambda) \) we have to smooth it over neighbouring frequencies. Now we impose the following assumptions on \( W(\cdot) \) and \( h(\cdot) \).

**Assumption 3.** The weighted function \( W : \mathbb{R} \rightarrow [0, \infty] \) satisfies \( W(0) = 0 \) for \( x \notin [-1/2, 1/2] \), and is a continuous and even function satisfying \( \int_{-1/2}^{1/2} W(x) dx = 1 \) and \( \int_{-1/2}^{1/2} x^2 W(x) dx < \infty \).

**Assumption 4.** The data taper \( h : \mathbb{R} \rightarrow \mathbb{R} \) satisfies (i) \( h(x) = 0 \) for all \( x \notin [0, 1] \) and \( h(x) = h(1-x) \), (ii) \( h(x) \) is continuous on \( \mathbb{R} \), twice differentiable at all \( x \notin \mathbb{R} \) where \( \mathbb{R} \) is a finite set of \( \mathbb{R} \), and \( \sup_{x \notin \mathbb{R}} \left| h''(x) \right| < \infty \). Write

\[
K_t(x) := \left\{ \int_{0}^{1} h(x)^2 dx \right\}^{-1} h(x + 1/2)^2, \quad x \in [-1/2, 1/2],
\]

which plays a role of kernel in the time domain.

Furthermore, we assume

**Assumption 5.** \( M = M(T) \) and \( N = N(T) \), \( M \ll N \ll T \) satisfy

\[
\frac{\sqrt{T}}{M^2} = o(1), \quad \frac{N^2}{T^2} = o(1), \quad \frac{\sqrt{T} \log N}{N} = o(1).
\]

(12)

The following lemmas are multivariate version of Theorem 2.2 of Dahlhaus (1996c) and Theorem A.2 of Dahlhaus (1997) (See also Sakiyama and Taniguchi (2003)).

**Lemma 1.** Assume that Assumptions 1-5 hold. Then

(i)

\[
E(I_{N}(u, \lambda)) = f(u, \lambda) + \frac{N^2}{2T^2} \int_{-1/2}^{1/2} x^2 K_t(x)^2 dx \frac{\partial^2}{\partial u^2} f(u, \lambda) + o\left( \frac{N^2}{T^2} \right) + O\left( \frac{\log N}{N} \right),
\]

(13)

(ii)

\[
E(\hat{f}(u, \lambda)) = f(u, \lambda) + \frac{N^2}{2T^2} \int_{-1/2}^{1/2} x^2 K_t(x)^2 dx \frac{\partial^2}{\partial u^2} f(u, \lambda) + \frac{1}{2M^2} \int_{-1/2}^{1/2} x^2 W(x)^2 dx \frac{\partial^2}{\partial \lambda^2} f(u, \lambda) + o\left( \frac{N^2}{T^2 + M^{-2}} \right) + O\left( \frac{\log N}{N} \right),
\]

(14)
where
\[ \| A \| \] is the Euclidean norm of the matrix \( A; \| A \| = \{ \text{tr}(AA^*) \}^{1/2}. \]

**Lemma 2.** Assume that Assumptions 1-5 hold. Let \( \phi_j(u, \lambda), j = 1, \ldots, k \) be \( m \times m \) matrix-valued continuous function on \([0, 1] \times [-\pi, \pi]\) which satisfies the same conditions as the transfer function matrix \( A(u, \lambda) \) in Assumption 2 and \( \phi_j(u, \lambda)^* = \phi_j(u, \lambda), \phi_j(u, -\lambda) = \phi_j(u, \lambda)^* \). Then

\[
L_T(\phi_j) = \sqrt{T} \left\{ \frac{1}{T} \sum_{t=1}^{T} \int_{-\pi}^{\pi} \text{tr} \left\{ \phi_j \left( \frac{t}{T}, \lambda \right) I_N \left( \frac{t}{T}, \lambda \right) \right\} d\lambda ight. \\
- \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr} \left\{ \phi_j (u, \lambda) f (u, \lambda) \right\} d\lambda du \right\}, \quad j = 1, \ldots, k
\]

have, asymptotically, a normal distribution with zero mean vector and covariance matrix \( V \) whose \((i, j)\)-element is

\[
4\pi \int_{0}^{1} \left[ \int_{-\pi}^{\pi} \text{tr} \{ \phi_i(u, \lambda) f(u, \lambda) \phi_j(u, \lambda) f(u, \lambda) \} d\lambda \right. \\
+ \frac{1}{4\pi^2} \sum_{a_1, a_2} \sum_{a_4} \kappa_{b_1, b_2, b_3, b_4} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_i(u, \lambda)_{a_1, a_2} \phi_j(u, \mu)_{a_4, a_3} \\
\left. \cdot A(u, \lambda)_{a_2, b_1} A(u, -\lambda)_{a_1, b_2} A(u, -\mu)_{a_3, b_4} A(u, \mu)_{a_4, b_3} d\lambda d\mu \right] du.
\]

Assumption 5 does not coincide with Assumption A.1 (ii) of Dahlhaus (1997). As mentioned in A.3 Remarks of Dahlhaus (1997), Assumption A.1 (ii) of Dahlhaus (1997) is required because of the \( \sqrt{T} \)-unbiasedness at the boundary 0 and 1. If we assume that \( \{X_{2-N/2, T}, \ldots, X_{0, T}\} \) and \( \{X_{T+1, T}, \ldots, X_{T+N/2, T}\} \) are available with Assumption 5, then from Lemma 1 (i)

\[
E (L_T(\phi_j)) = \sqrt{T} E \left\{ \frac{1}{T} \sum_{t=1}^{T} \int_{-\pi}^{\pi} \text{tr} \left\{ \phi_j \left( \frac{t}{T}, \lambda \right) I_N \left( \frac{t}{T}, \lambda \right) \right\} d\lambda \\
- \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr} \left\{ \phi_j (u, \lambda) f (u, \lambda) \right\} d\lambda du \right\} \\
= O \left( \sqrt{T} \left( \frac{N^2}{T^2} + \frac{\log N}{N} + \frac{1}{T} \right) \right) = o(1).
\]
3. Measures of disparity. We suppose that we have a collection of zero-mean $m$-dimensional vector locally stationary time series $X_{i,T} = (X_{i,T}^{(1)}, X_{i,T}^{(2)}, \ldots, X_{i,T}^{(m)})', t = 1, 2, \ldots, T$. Denote by $p_i(x)$, $i = 1, 2$, the probability density functions of the $mT \times 1$ vector $x = (X_{1,T}', X_{2,T}', \ldots, X_{T,T}')'$ under two hypotheses $\Pi_i$, $i = 1, 2$, respectively. In the case of locally stationary processes, $\Pi_i$, $i = 1, 2$ are, respectively, described by the time varying spectral density matrices $f_i(u, \lambda)$, $i = 1, 2$ corresponding to $mT \times mT$ matrices $\Sigma_T(p_i)$, $i = 1, 2$. Although theory developed later transcends the usual normal theory, it is convenient to use the normal assumption temporarily to motivate measures of disparity between the densities $p_i(\cdot)$, $i = 1, 2$.

One classical measure of disparity between two multivariate densities is the Kullback-Leibler (KL) discrimination information, defined by

$$K(p_j; p_k) = E_p \left\{ \log \frac{p_j(x)}{p_k(x)} \right\},$$

where $E_p$ denotes the expectation under the density $p(\cdot)$. The KL discrimination information takes the form

$$K(p_j; p_k) = \frac{1}{2} \left( \text{tr}\{\Sigma_T(p_j)\Sigma_T^{-1}(p_k)\} - \log \left| \frac{\Sigma_T(p_j)}{\Sigma_T(p_k)} \right| - mT \right)$$

when $p_i(x)$ correspond to two hypothetical zero-mean multivariate normal distributions. The $mT \times mT$ covariance matrices $\Sigma_T(p_i)$ contain the $m \times m$ matrices $\Sigma_{s,t}^T(p_i)$, $s, t = 1, \ldots, T$ as blocks, where

$$\Sigma_{s,t}^T(p_i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i \lambda(s - t)) A_{s,t}^\alpha(\lambda) \Omega A_{s,t}^\alpha(-\lambda)' d\lambda.$$

Parzen (1990) proposed to use the Chernoff (CH) information measure

$$B_\alpha(p_j; p_k) = -\log E_{p_j} \left\{ \left( \frac{p_j(x)}{p_k(x)} \right)^\alpha \right\},$$

as a measure of disparity between the two densities, where the measure is indexed by $\alpha$, $0 < \alpha < 1$. For $\alpha = \frac{1}{2}$, the Chernoff information measure is the symmetric divergence measure. For two normal random vectors differing only in the covariance structure, the measure (23) takes the form

$$B_\alpha(p_j; p_k) = \frac{1}{2} \left( \log \left| \frac{\alpha \Sigma_T(p_j) + (1 - \alpha) \Sigma_T(p_k)}{\Sigma_T(p_k)} \right| - \alpha \log \left| \frac{\Sigma_T(p_j)}{\Sigma_T(p_k)} \right| \right).$$

It is of interest to note the antisymmetry property $B_\alpha(p_j; p_k) = B_1-\alpha(p_k; p_j)$ and that $B_\alpha(p_j; p_k)$, scaled by $\alpha(1 - \alpha)$ converges to $K(p_j; p_k)$ for $\alpha \to 0$ and to $K(p_k; p_j)$ for $\alpha \to 1$. Hence the Chernoff measure behaves like the two Kullback-Leibler measures for values of the parameter $\alpha$ that are near the boundaries 0 and 1.

It should be recognized that the information measures (21) and (24) both involve $mT \times mT$ matrices whose dimensions tend to infinity as $T \to \infty$. As in the case of stationary processes, it is natural to use spectral approximations in terms of the time varying spectral density matrices $f_i(u, \lambda)$, $i = 1, 2$. The appropriate versions of (21) and (24) are

$$K(f_j; f_k) = \lim_{T \to \infty} T^{-1} K(p_j; p_k)$$

$$= \frac{1}{2} \int_0^1 \int_{-\pi}^\pi \left( \text{tr}\{f_j(u, \lambda)f_k^{-1}(u, \lambda)\} - \log \left| \frac{f_j(u, \lambda)}{f_k(u, \lambda)} \right| - m \right) \frac{d\lambda}{2\pi} d\mu$$
and

\[ B_\alpha(f_j; f_k) = \lim_{T \to \infty} T^{-1} B_\alpha(p_j; p_k) \]

(26) \[ = \frac{1}{2} \int_0^1 \int_{-\pi}^{\pi} \left( \log \frac{|\alpha f_j(u, \lambda) + (1 - \alpha) f_k(u, \lambda)|}{|f_k(u, \lambda)|} \right) - \alpha \log \frac{|f_j(u, \lambda)|}{|f_k(u, \lambda)|} \frac{d\lambda}{2\pi} du. \]

Note here that the time-varying spectral matrices \( f_j(u, \lambda) \) correspond to the multivariate densities \( p_j(x) \). The advantage of (25) and (26) is that the evaluation problem is reduced to inverting \( m \times m \) matrices. Both forms (25) and (26) are functions of the matrix product \( f_j(u, \lambda) f_k^{-1}(u, \lambda) \) and can be generalized to the following disparity measure

(27) \[ D_H(f_j; f_k) = \frac{1}{2} \int_0^1 \int_{-\pi}^{\pi} H(f_j(u, \lambda) f_k^{-1}(u, \lambda)) \frac{d\lambda}{2\pi} du \]

where \( H(\cdot) \) is some matrix-valued function. To ensure that \( D_H(f_j; f_k) \) has the quasi-distance property, we require \( D_H(f_j; f_k) \geq 0 \), and that the equality holds if and only if \( f_j = f_k \) almost everywhere. The function \( H(Z) \) must have a unique minimum at \( Z = E_m \), the identity matrix. There are many possible choices of \( H(Z) \) such that \( D_H(\cdot; \cdot) \) satisfies the quasi-distance property, but we consider only the two cases corresponding to (25) and (26):

(28) \[ H_K(Z) = \text{tr} \{ Z \} - \log |Z| - m \]

and

(29) \[ H_B(\alpha, Z) = \log |\alpha Z + (1 - \alpha)E_m| - \alpha \log |Z|. \]

Note that another possible choice is the quadratic function

(30) \[ H_Q(Z) = \frac{1}{2} \text{tr}(Z - E_m)^2. \]

Generally, \( D_H(\cdot; \cdot) \) is not symmetric but can easily be made so by defining

(31) \[ \tilde{H}(Z) = H(Z) + H(Z^{-1}). \]

The general form (27) can be approximated by sums over frequencies of the form \( \lambda_s = 2\pi s/T \) and time \( \{u_t = t/T\} \), \( s, t = 1, 2, \ldots, T \), i.e.,

(32) \[ D_H(f_j; f_k) \approx \frac{1}{2} T^{-2} \sum_{s,t=1}^T H(f_j(u_t, \lambda_s) f_k^{-1}(u_t, \lambda_s)). \]

4. Discriminant analysis. Suppose that we wish to investigate the problem of classifying a realization \( X_T = (X_{t-N/2,T}, \ldots, X_{1,T}, \ldots, X_{T,T}, \ldots, X_{T+N/2,T}) \) into one of two known categories \( \Pi_j, j = 1, 2 \), where \( \Pi_j \) is described by the time varying spectral density matrix \( f_j(u, \lambda) \). Let \( \hat{f}_T(u, \lambda) \) be the nonparametric time varying spectral density estimator given by (9), which is based on observation to be classified. We measure the disparity between the sample spectrum of \( X_T \) and category \( \Pi_j \) by \( D_H(\hat{f}_T; f_j) \). Then the proposed rule is to classify \( X_T \) into \( \Pi_1 \) or \( \Pi_2 \) according to \( D_H > 0 \) or \( D_H \leq 0 \), where \( D_H \) is the discriminant function defined by

(33) \[ D_H = D_H(\hat{f}_T; f_2) - D_H(\hat{f}_T; f_1). \]

In this section we examine the asymptotic properties of discriminant function (33). Assume that the category \( \Pi_j \) is an \( m \)-variate linear process of the form \( X_{i,T} = \sum_{k=-\infty}^{\infty} a_{i,T}(k) \xi_{t-k}, \)
where \( m \times m \) matrices \( a_{j,t}^{(j)}(k) \)'s and i.i.d. \( m \times 1 \) zero mean vectors \( \varepsilon_t \)'s satisfy Assumptions 1 and 2. The use of \( \hat{f}_T(u, \lambda) \) instead of the data tapered periodogram \( I_N(u, \lambda) \) is essential, because \( D_H(I_N; g) \) does not converge in probability to \( D_H(f_j; g) \) under \( \Pi_j \) if \( D_H(I_N; g) \) is nonlinear with respect to \( I_N \) (See Taniguchi and Kakizawa (2000)).

We discuss the performance of the discriminant function (33). First, we evaluate the asymptotics of the misclassification probabilities based on \( D_H \):

\[
P_{D_H}(2|1) = \Pr(D_H \leq 0|\Pi_1)
\]

and

\[
P_{D_H}(1|2) = \Pr(D_H > 0|\Pi_2).
\]

It is shown that \( P_{D_H}(2|1) \) and \( P_{D_H}(1|2) \) converge to zero as \( T \to \infty \) if \( f_1(u, \lambda) \neq f_2(u, \lambda) \).

Next assuming that \( \Pi_1 \) is contiguous to \( \pi_2 \), the limit of the two misclassification probabilities is evaluated. Then we will elucidate the asymptotic optimality and robustness.

In what follows, set \( (j, k) = (1, 2) \) or \( (2, 1) \). We give the following assumption on matrix-valued function \( H(Z) \).

**Assumption 6.** (i) \( H : \mathbb{C}^{m \times m} \to \mathbb{R} \) is real-valued holomorphic function defined on an open set \( D \) in \( \mathbb{C}^{m \times m} \).

(ii) \( H(E_m) = 0 \) and \( H^{(1)}(E_m) = 0 \) \( (m \times m \) zero matrix), where \( H^{(1)}(\{\cdot\}) \) is the first derivative of \( Z \) at \( \{\cdot\} \) whose \((a, b)\)-th element is \( \frac{\partial}{\partial z_{ab}} H(Z) \). The \( m^2 \times m^2 \) Hessian matrix of \( H(Z) \), defined by

\[
\frac{\partial}{\partial (\text{vec } Z)'} \left( \frac{\partial}{\partial (\text{vec } Z)'} H(Z) \right)'
\]

is positive definite at \( Z = E_m \). That is, \( H(Z) \) has a unique minimum zero at \( Z = E_m \).

(iii) The \( m \times m \) matrix \( Q_{j,k}(u, \lambda) \) defined by

\[
Q_{j,k}(u, \lambda) = f_k^{-1}(u, \lambda) \left[ H^{(1)}(f_j(u, \lambda)f_k^{-1}(u, \lambda)) \right]'
\]

satisfies \( Q_{j,k}(u, \lambda)^* = Q_{j,k}(u, \lambda) \) and \( Q_{j,k}(u, -\lambda) = Q_{j,k}(u, \lambda)' \).

**Theorem 1.** Under the Assumptions 1-6, suppose that \( f_1(u, \lambda) \neq f_2(u, \lambda) \) on a set of positive Lebesgue measure. Then under \( \Pi_j \), \( D_H \xrightarrow{D} (-1)^{j+1} D_H(f_j; f_k) \) and

\[
\sqrt{T} \{ D_H + (-1)^j D_H(f_j; f_k) \} \xrightarrow{D} N(0, V_H^2(j, k)), \quad \text{as } T \to \infty,
\]

where \( D_H(f_j; f_k) \) is the integral disparity (27) and

\[
V_H^2(j, k) = \int_0^1 \left[ \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr}(Q_{j,k}(u, \lambda)f_j(u, \lambda))^2 d\lambda 
+ \frac{1}{64\pi^5} \sum_{a,b,c,d} \kappa_{a,b,c,d}(\gamma_{a,b}^{(j,k)}(u)\gamma_{c,d}^{(j,k)}(u)) \right] du,
\]

with

\[
\Gamma_H^{(j,k)}(u) = \left\{ \gamma_{a,b}^{(j,k)}(u) \right\}_{a,b=1,\ldots,m} = \int_{-\pi}^{\pi} A_s(u, \lambda)Q_{j,k}(u, \lambda)A_k(u, \lambda) d\lambda.
\]
In view of Theorem 1, the limiting forms of misclassification probabilities (34) and (35) satisfy
\[ \lim_{T \to \infty} P_{DH}(k|j) = 0 \] for \((j, k) = (1, 2), (2, 1)\). This shows that the discriminant
\(D_H\) is consistent. From (39), one may also approximate them as the normal integrals
\[ P_{DH}(k|j) \approx \Phi \left( -\sqrt{T} \frac{D_H(f_j; f_k)}{V_H(j, k)} \right), \]
which depend on the fourth-order cumulants unless (40) is a zero matrix. To look at
robustness, we assume that the hypothetical \(m\)-variate linear process is generated by
\[ X_{t,T} = \sum_{s=\infty}^{\infty} a^{(1)}_{t,T}(s) \varepsilon_{t-s} \]
under \(\Pi_1\) and by
\[ X_{t,T} = \sum_{s=\infty}^{\infty} \left\{ a^{(1)}_{t,T}(s) + T^{-1/2} a^{(2)}_{t,T}(s) \right\} \varepsilon_{t-s} \]
under \(\Pi_2\). Thus, the time varying spectral densities associated with \(\Pi_1\) and \(\Pi_2\) are
\[ f_1(u, \lambda) = A^{(1)}(u, \lambda) \Omega A^{(1)}(u, \lambda)^* \]
and
\[ f_2(u, \lambda) = \left\{ A^{(1)}(u, \lambda) + T^{-1/2} A^{(2)}(u, \lambda) \right\} \Omega \]
\[ = \left\{ A^{(1)}(u, \lambda) + T^{-1/2} A^{(2)}(u, \lambda) \right\}^*, \]
with \(A^{(i)}(u, \lambda) = \sum_{s=\infty}^{\infty} a^{(i)}_{s}(u) \exp\{-i\lambda s\}, i = 1, 2\). The quantities \(D_H(f_j; f_k)\) and \(V_H(j, k)\) are determined by the local property of the function \(H(Z)\) at \(Z = E_m\).

**Assumption 7.** The \(m^2 \times m^2\) Hessian matrix of \(H(Z)\) at \(E_m\) is \(cK_m\), where \(K_m\) is the commutation matrix (e.g., Magnus and Neudecker (1988)) and \(c > 0\).

Note that \(H_K\), \(H_{B_n}\) and \(H_Q\) in (28), (29) and (30) satisfy Assumptions 6 and 7.

**Theorem 2.** Let \(f_1\) and \(f_2\), defined by (44) and (45), be the hypothetical time varying hypothesis spectral density matrices of \(m\)-variate linear processes (42) and (43), respectively. Under Assumptions 1-7, if \(\Theta_0^{(2)}(u) = 0\) \((m \times m\) zero matrix\), the asymptotic misclassification probabilities are independent of the non-Gaussianity of the process, and are given by
\[ \lim_{T \to \infty} P_{DH}(2|1) = \lim_{T \to \infty} P_{DH}(1|2) \]
\[ = \Phi \left( \frac{1}{2} \sqrt{\frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \text{tr}\{\Delta(u, \lambda)\}^2 d\lambda du} \right), \]
with
\[ \Delta(u, \lambda) = \left\{ A^{(1)}(u, \lambda) \Omega A^{(2)}(u, \lambda)^* + A^{(2)}(u, \lambda) \Omega A^{(1)}(u, \lambda)^* \right\} \]
\[ \{A^{(1)}(u, \lambda) \Omega A^{(1)}(u, \lambda)^* \}^{-1}. \]
Furthermore, if the process concerned is Gaussian, the exact Gaussian log-likelihood ratio is

$$\Lambda_T(p_1, p_2) \equiv \frac{1}{T} \text{ Gaussian log likelihood ratio}$$

$$= \frac{1}{2T} X_T' (\Sigma^T(p_1)^{-1} - \Sigma^T(p_2)^{-1}) X_T - \frac{1}{2T} \log \frac{|\Sigma^T(p_2)|}{|\Sigma^T(p_1)|}.$$  

According to Proposition 2.5 and Lemma A.8 of Dahlhaus (2000), it is seen that, under $\Pi_j$, for each $\varepsilon > 0$,

$$E \left( \sqrt{T} \Lambda_T(p_1, p_2) \right)$$

$$= \frac{\sqrt{T}}{2T} \left\{ tr \left\{ \Sigma^T(p_j) (\Sigma^T(p_1)^{-1} - \Sigma^T(p_2)^{-1}) \right\} \right. - \log \left| \Sigma^T(p_2) \right| \left| \Sigma^T(p_1) \right|$$

$$+ O(T^{-\frac{1}{2}+\varepsilon} + T^{-\frac{1}{4}} \log^{11} T)$$

$$= (-1)^{j+1} \sqrt{T} D_{\Delta}(\mathbf{f}_j; \mathbf{f}_k) + o(1)$$

and

$$Var \left( \sqrt{T} \Lambda_T(p_1, p_2) \right) = \frac{1}{2T} tr \left\{ \Sigma^T(p_j) (\Sigma^T(p_1)^{-1} - \Sigma^T(p_2)^{-1}) \right\}^2$$

$$= \frac{1}{4T} \left\{ \int_{0}^{1} \int_{\pi}^{\pi} \left( \left\{ \mathbf{f}_j(u, \lambda) (\mathbf{f}_2^{-1}(u, \lambda) - \mathbf{f}_1^{-1}(u, \lambda)) \right\} \right)^2 d\lambda d\mu \right\}^2$$

$$+ O(T^{-1} \log^{23} T).$$

Therefore,

$$\lim_{T \to \infty} P_{GLR}(2|1) = \lim_{T \to \infty} P_{GLR}(1|2)$$

$$= \lim_{T \to \infty} P_{DH}(2|1) = \lim_{T \to \infty} P_{DH}(1|2)$$

$$= \Phi \left( -\frac{1}{2} \sqrt{\frac{1}{4T} \int_{0}^{1} \int_{\pi}^{\pi} \{ \mathbf{A}(u, \lambda) \}^2 d\lambda d\mu \right),$$

where

$$P_{GLR}(2|1) = Pr(\Lambda_T(p_1, p_2) \leq 0|\Pi_1)$$

and

$$P_{GLR}(1|2) = Pr(\Lambda_T(p_1, p_2) > 0|\Pi_2),$$

that is, the discriminant criterion based on $D_H$ is asymptotically Gaussian optimal.

**Remark 1. Peak Robustness of $B_\alpha$.**

We consider a case where the time varying spectral density of $X_T$ is contaminated by a sharp peak. In this case, we can see that $B_\alpha(\mathbf{f}_j; \mathbf{f}_k)$ is robust with respect to peak, but $K(\mathbf{f}_j; \mathbf{f}_k)$ is not so. Define

$$\tilde{f}_i(u, \lambda) = \begin{cases} 
\mathbf{f}_i(u, \lambda) & \text{on } \Omega = [-\pi, \pi] - \Omega_\varepsilon; \\
\mathbf{f}_i(u, \lambda)/\varepsilon^r & \text{on } \Omega_\varepsilon,
\end{cases}$$

(54)
where $\Omega = [\lambda_0, \lambda_0 + \epsilon]$ is an interval in $[-\pi, \pi]$ for sufficiently small $\epsilon > 0$ and $r > 1$. Suppose that $f_j(u, \lambda) \neq f_k(u, \lambda)$ on a set of a positive Lebesgue measure. Then, under Assumption 2, it can be shown that

$$\lim_{\epsilon \to 0} |B_\alpha(\bar{f}, f_j, f_k) + (-1)^j B_\alpha(f_j, f_k)| = 0 \quad \text{for} \quad \alpha \in (0, 1),$$

where

$$\lim_{\epsilon \to 0} |K(\bar{f}, f_j, f_k) + (-1)^j K(f_j, f_k)| = \infty$$

and

$$B_\alpha(\bar{f}, f_j, f_k) = B_\alpha(f_j, f_k) - B_\alpha(f_k, f_j)$$

and

$$K(\bar{f}, f_j, f_k) = K(f_j, f_k) - K(\bar{f}, f_j)$$

That is, $B_\alpha(\bar{f}; f_j, f_k)$ is insensitive to sharp peak in the spectral density, while $K(f_j; f_k)$ is sensitive. Thus, the discriminant statistic $B_\alpha(\bar{f}; f_j, f_k)$ is better than $K(f_j; f_k)$ if the time varying spectral density of $X_T$ is contaminated by a sharp peak.

**Numerical example.** We consider the time varying AR(2) model

$$X_{t,T} = b_0^{(1)} \left( \frac{t}{T} \right) X_{t-1,T} + b_0^{(2)} \left( \frac{t}{T} \right) X_{t-2,T} + \varepsilon_t,$$

with $b_0^{(i)}(u) = \theta^{(i)} \exp(-2\theta^{(i)}(u + 1))$, $i = 1, 2$, and $\varepsilon_t$ are i.i.d. random variables with probability density

$$p(x) = \exp(-(x + 1)), \quad x > -1.$$

Then the time varying spectral density is given by

$$f_0(u, \lambda) = \frac{1}{2\pi} \left| \frac{1}{1 - \theta^{(1)} \exp(-2\theta^{(1)}(u + 1)) \exp(i\lambda)} \right|^2.$$

Now, we give two examples for the model (59):

**Example 1** $(\theta^{(1)}, \theta^{(2)}) = (0.8, 0)$: AR(1) model.

**Example 2** $(\theta^{(1)}, \theta^{(2)}) = (0.8, 0.2)$: AR(2) model.

In Figure 1, AR(1) coefficient $b_0^{(1)}(u)$ (real line) and AR(2) coefficient $b_0^{(2)}(u)$ (dotted line) of Example 2 are plotted. From this figure, it is seen that $b_0^{(1)}(u)$ decreases as $u \to 1$, on the other hand $b_0^{(2)}(u)$ is almost constant.

**Figures 1 is about here.**

For Examples 1 and 2, the observed stretch $\{X_{t,T}\}$ with $T = 2^8$ are plotted in Figures 2 and 3, respectively. The time varying spectral densities of them are, respectively, given in Figures 4 and 5.
Figures 2 and 3 are about here.

Figures 4 and 5 are about here.

Next, in Example 1, let the time varying spectral densities associated with $\Pi_1$ and $\Pi_2$ be

$$
\begin{cases}
 \Pi_1 : f_1(u, \lambda) = \frac{1}{2\pi} \left| \frac{1 - \theta_1^{(1)} \exp(-2\theta_1^{(1)}(u+1)) \exp(i\lambda)}{1 - \theta_1^{(1)} \exp(-2\theta_1^{(1)}(u+1)) \exp(i\lambda)} \right|^2, \\
 \Pi_2 : f_2(u, \lambda) = \frac{1}{2\pi} \left| \frac{1 - \theta_2^{(1)} \exp(-2\theta_2^{(1)}(u+1)) \exp(i\lambda)}{1 - \theta_2^{(1)} \exp(-2\theta_2^{(1)}(u+1)) \exp(i\lambda)} \right|^2,
\end{cases}
$$

respectively. We consider the following two cases;

CASE I The time varying spectral density of process concerned is not contaminated by a sharp peak, That is, we actually observe

$$Y_{t,T} = X_{t,T}.$$

CASE II The time varying spectral density of process concerned is contaminated by a sharp peak. That is, we actually observe

$$Y_{t,T} = X_{t,T}^1 + X_{t,T}^2,$$

where

$$X_{t,T}^1 = X_{t,T}$$

and

$$X_{t,T}^2 = \int_{-\pi}^{\pi} \left\{ e^{-r/2} - 1 \right\} A_0 \left( \frac{t}{T}, \lambda \right) \exp(i\lambda t) I(\lambda_0, \lambda_0 + \epsilon) d\lambda.$$

For the cases II, the observed stretch $\{Y_{t,T}\}$ with $T = 2^k$, $\theta^{(1)} = 0.8$, $\lambda_0 = -\pi/4$, $r = 3$ and $\epsilon = 0.01$ are plotted in Figures 6. The time varying spectral density $f_1(u, \lambda)$ are given in Figures 7.

Figures 6 and 7 are about here.

Finally, in Figures 8 and 9, we plot the pair graphs $K(\bar{f}_1, f_1, f_2)$, $K(f_1, f_2)$ and $B_{1/2}(\bar{f}_1, f_1, f_2)$, $B_{1/2}(f_1, f_2)$ with $(\theta_1^{(1)}, \theta_2^{(1)}) = (0.8, 0.9)$, respectively. From Figures 8 and 9, it is seen that $K(\bar{f}_1, f_1, f_2)$ diverges as $\epsilon \to 0$, on the other hand $B_{1/2}(f_1, f_2)$ converges to $B_{1/2}(f_1, f_2)$ as $\epsilon$ tends to 0.

Figures 8 and 9 are about here.
5. Asymptotically non-Gaussian optimal classification. Suppose that $X_T = (X_{1,T}, \ldots, X_{T,T})'$ is a realization of a scalar-valued locally stationary process with transfer function $A_\theta^0$ where the corresponding $A_\theta$ is uniformly bounded from above and below, and its time varying spectral density $f_\theta(u, \lambda) := |A_\theta(u, \lambda)|^2$ depends on a parameter vector $\theta = (\theta_1, \ldots, \theta_q) \in \Theta \subset \mathbb{R}^q$.

Let $\Pi_1$ and $\Pi_2$ be two categories with probability density functions $p_1(X)$ and $p_2(X)$, respectively. We investigate the problems of classifying a locally stationary process $\{X_T\}$ into one of two categories described by two hypotheses:

$$
\begin{align*}
\Pi_1 : f_1(u, \lambda) &= f(u, \lambda|\theta) \\
\Pi_2 : f_2(u, \lambda) &= f(u, \lambda|\theta_T + \frac{h}{\sqrt{T}}),
\end{align*}
$$

where $\theta \in \Theta \subset \mathbb{R}^q$ and $h = (h_1, \ldots, h_q)'$. We assign the observed stretch $X_T$ to category $\Pi_1$ if $X_T$ falls into region $R_1$; otherwise we assign it to $\Pi_2$, where $R_1$ and $R_2$ are exclusive and exhaustive regions in $\mathbb{R}^q$. It is well known that the classification regions defined by

$$
R_1 = \left[ X_T : A_T(p_1, p_2) = \log \frac{p_1(X_T)}{p_2(X_T)} > 0 \right]
$$

give the optimal classification (See Anderson (1984)).

Introducing the notations $\nabla_\theta = \frac{\partial}{\partial \theta}$, $\nabla = (\nabla_1, \ldots, \nabla_q)'$, $\nabla_{ij} = \frac{\partial}{\partial \theta_i \theta_j}$, $\nabla^2 = (\nabla_{ij})_{i,j=1,\ldots,q}$, we make the following assumption.

**Assumption 8. (A1)** There exists a constant $K$ with

$$
\sup_{t,\lambda} \left| \nabla^s \left[ A_{\theta,t,T}^0 - A_\theta \left( \frac{t}{T}, \lambda \right) \right] \right| \leq KT^{-1}
$$

for $s = 0, 1, 2$. The components of $A_\theta(u, \lambda)$, $\nabla A_\theta(u, \lambda)$ and $\nabla^2 A_\theta(u, \lambda)$ are differentiable in $u$ and $\lambda$ with uniformly continuous derivatives $\frac{\partial}{\partial u \partial \lambda}$.

Writing

$$
\varepsilon_t = \int_{-\pi}^{\pi} \exp (i\lambda t) d\xi(\lambda),
$$

we assume the followings.

**Assumption 9. (B1)** $\varepsilon_t$’s are i.i.d. random variables with mean 0, variance 1 and finite fourth order moment $E(\varepsilon_t^4)$. Furthermore the distribution is absolutely continuous with respect to Lebesgue measure, and has the probability density $p(z) > 0$ on $\mathbb{R}$.

(B2) $p(\cdot)$ satisfies

$$
\lim_{|z| \to \infty} p(z) = 0, \quad \text{and} \quad \lim_{|z| \to \infty} |zp(z)| = 0.
$$

(B3) The continuous derivatives $Dp$, $D^2p \equiv D(Dp)$ and $D^3p \equiv D(D^2p)$ of $p(\cdot)$ exist on $\mathbb{R}$, and $D^3p$ satisfies the Lipschitz condition.

(B4) $\mathcal{F}(p) = \int (\phi(z))^2 p(z) dz < \infty$.
\( (73) \int (D^s \phi(z))^2 p(z) dz < \infty, \ s = 1, 2, \)

\( (74) E(\varepsilon_t \phi^2(\varepsilon_t)), \ E(\varepsilon_t^2 \phi^2(\varepsilon_t)) \) and \( E(\phi(z))^4 < \infty \)

and

\( (75) \int D^2 p(z) dz = 0, \lim_{|z| \to \infty} Dp(z) z^2 = 0, \)

where \( \phi(z) = \frac{Dp(z)}{p(z)} \).

**Assumption 10. (C1) \{X_{\theta,T}\} has the MA(\infty) and AR(\infty) representations**

\( (76) X_{\theta,T} = \sum_{j=0}^{\infty} a_{\theta,T}(j) \varepsilon_{t-j}, \)

\( (77) a_{\theta,T}(0) \varepsilon_t = \sum_{k=0}^{\infty} b_{\theta,T}(k) X_{t-k,T}, \)

where \( a_{\theta,T}(j), \ b_{\theta,T}(k) \in \mathbb{R}, \ b_{\theta,T}(0) \equiv 1 \) and \( a_{\theta,T}(j) = a_{\theta,0,T}(j) = a_{\theta}(j) \) for \( t \leq 0. \)

(C2) Every \( a_{\theta,T}(j) \) is continuously three times differentiable with respect to \( \theta \), and the derivatives satisfy

\( (78) \sup_{t,T} \left\{ \sum_{j=0}^{\infty} (1 + j) |\nabla_i \cdots \nabla_s a_{\theta,T}(j)| \right\} < \infty \) for \( s = 0, 1, 2, 3. \)

(C3) Every \( b_{\theta,T}(k) \) is continuously three times differentiable with respect to \( \theta \), and the derivatives satisfy

\( (79) \sup_{t,T} \left\{ \sum_{k=0}^{\infty} (1 + k) |\nabla_i \cdots \nabla_s b_{\theta,T}(k)| \right\} < \infty \) for \( s = 0, 1, 2, 3. \)

(C4)

\( (80) a_{\theta,T}(0) = \exp \left\{ \frac{1}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f_{\theta,T}(\lambda)) d\lambda \right\}, \)

where \( f_{\theta,T}(\lambda) = |A_{\theta,T}(\lambda)|^2. \)

By (76) and (77) we have

\( (81) a_{\theta,T}(0) \varepsilon_t = \sum_{k=0}^{t-1} b_{\theta,T}(k) X_{t-k,T} + \sum_{r=0}^{\infty} c_{\theta,T}(r) \varepsilon_{t-r}, \)

where

\( (82) c_{\theta,T}(r) = \sum_{s=0}^{r} b_{\theta,T}(t+s) a_{\theta}(r-s). \)
From Assumption 10 it follows that
\[
\sum_{r=0}^{\infty} |c_{\theta,t,T}^r(r)| \leq \sum_{k=t}^{\infty} \sum_{l=0}^{\infty} |b_{\theta,t,T}^k(k)||a_{\theta}^l(l)|
\]
\[
\leq \sum_{l=0}^{\infty} |a_{\theta}^l(l)| \frac{1}{t} \sum_{k=t}^{\infty} k|b_{\theta,t,T}^k(k)| = O(t^{-1}).
\]
(83)

By Theorem 1 of Hirukawa and Taniguchi (2004), we have for all \( \theta \in \Theta \), under \( \Pi_1 \), as \( T \to \infty \), the log-likelihood ratio \( \Lambda_T(p_1, p_2) \) has, asymptotically, normal distribution \( \mathcal{N}(\frac{1}{2} h'\Gamma(\theta)h, h'\Gamma(\theta)h) \), where
\[
\Gamma(\theta) = \int_0^1 \left\{ \frac{\mathcal{F}(p)}{4\pi} \int_{-\pi}^{\pi} \frac{(\nabla f_\theta(u, \lambda))(\nabla f_\theta(u, \lambda))'}{|f_\theta(u, \lambda)|^2} d\lambda 
+ \frac{1}{16\pi^2} \left\{ E(\epsilon^s T^s) - 2\mathcal{F}(p) - 1 \right\} \right\} du.
\]
(84)

Furthermore, under \( \Pi_2 \), \( \Lambda_T(p_1, p_2) \overset{D}{\rightarrow} \mathcal{N}(\frac{1}{2} h'\Gamma(\theta)h, h'\Gamma(\theta)h) \), hence
\[
\lim_{T \to \infty} P_{LR}(2|1) = \lim_{T \to \infty} P_{LR}(1|2) = \Phi \left( -\frac{1}{2} \sqrt{h'\Gamma(\theta)h} \right),
\]
(85)

where
(86) \( P_{LR}(2|1) = Pr(\Lambda_T(p_1, p_2) \leq 0|\Pi_1) \)
and
(87) \( P_{LR}(1|2) = Pr(\Lambda_T(p_1, p_2) > 0|\Pi_2) \).

Since \( (\epsilon_s, s \leq 0) \) are unobservable, instead of \( \Lambda_T(p_1, p_2) \) we use the “quasi-log-likelihood ratio”
\[
F_T(p_1, p_2) = \log \frac{F_T(\theta)}{F_T(\theta T)},
\]
(88)

with
\[
F_T(\theta) = \prod_{t=1}^{T} \frac{1}{a_{\theta,t,T}^0(0)} p \left\{ \sum_{k=0}^{t-1} b_{\theta,t,T}^k(k)X_{t-k,T} \right\},
\]
(89)

for classification criterion.

**Theorem 3.** The discriminant criterion based on the quasi-log-likelihood ratio is asymptotically optimal.

Since hypotheses are often unknown, when there are training samples, the results may also be extended to the plug-in version of \( D_H \), obtained by replacing \( f_j \) with the estimator \( \bar{f}_j \) based on the training samples.

6. **Proofs.** This section provides the proofs of theorems.
Proof of Theorem 1. Let
\[
\hat{H}_j(u, \lambda) = H \left( \hat{f}_T(u, \lambda) \hat{f}_j^{-1}(u, \lambda) \right) - H \left( f_j(u, \lambda) f_j^{-1}(u, \lambda) \right)
\]
(90)
then from Lemma 1, the same argument as in Theorem 1 of Taniguchi et al. (1996), leads to, under \( \Pi_j \)
\[
\hat{H}_j(u, \lambda) = O_P \left( \frac{M}{N} \right)
\]
(91)
and
\[
H \left( \hat{f}_T(u, \lambda) \hat{f}_j^{-1}(u, \lambda) \right) = O_P \left( \frac{M}{N} \right).
\]
(92)
uniformly in \( \lambda \) and \( u \). Since, \( D_H \) is written as
\[
D_H = \frac{1}{4\pi} \int_0^L \int_{-\pi}^\pi H \left( \hat{f}_T(u, \lambda) \hat{f}_j^{-1}(u, \lambda) \right) - H \left( f_j(u, \lambda) f_j^{-1}(u, \lambda) \right) d\lambda du,
\]
(93)
From (91) and (92), under \( \Pi_j \)
\[
\sqrt{T} \left\{ D_H + (-1)^j D_H (f_j; f_k) \right\}
\]
\[
= \frac{(-1)^{j+1} \sqrt{T}}{4\pi} \int_0^L \int_{-\pi}^\pi tr \left\{ Q_{j,k}(u, \lambda) (\hat{f}_T(u, \lambda) - f_j(u, \lambda)) \right\} d\lambda du + o_P(1)
\]
(94)
According to Lemma 2,
\[
L_T \left( \frac{(-1)^{j+1}}{4\pi} Q_{j,k} \right)
\]
\[
= \frac{(-1)^{j+1} \sqrt{T}}{4\pi} \int_0^L \int_{-\pi}^\pi tr \left\{ Q_{j,k}(u, \lambda) (I_N(u, \lambda) - f_j(u, \lambda)) \right\} d\lambda du + O(T^{-\frac{1}{2}})
\]
(95)
have, asymptotically, a normal distribution with zero mean vector and covariance matrix \( V_H^2(j, k) \). Thus, the proof of Theorem 1 is complete if we show \( S_T - L_T = o_P(1) \). From the definition of \( f \), it follows that
\[
S_T - L_T
\]
\[
= \sqrt{T} \int_0^L \int_{-\pi}^\pi tr \left\{ Q_{j,k}(u, \lambda) \left( \int_{-\pi}^\pi W_T(\lambda - \mu) f_j(u, \mu) d\mu - f_j(u, \lambda) \right) \right\} d\lambda du
\]
\[
+ \sqrt{T} \int_0^L \int_{-\pi}^\pi tr \left\{ Q_{j,k}(u, \lambda) \int_{-\pi}^\pi (I_N(u, \mu) - f_j(u, \mu)) W_T(\lambda - \mu) d\mu \right\} d\lambda du
\]
\[
- \sqrt{T} \int_0^L \int_{-\pi}^\pi tr \left\{ Q_{j,k}(u, \lambda) (I_N(u, \mu) - f_j(u, \mu)) \right\} d\lambda du
\]
\[
= \sqrt{T} \int_0^L \int_{-\pi}^\pi tr \left\{ Q_{j,k}(u, \lambda) \left( \int_{-\pi}^\pi W_T(\lambda - \mu) f_j(u, \mu) d\mu - f_j(u, \lambda) \right) \right\} d\lambda du
\]
\[ + \sqrt{T} \int_0^1 \int_{-\pi}^\pi tr \{ D(u, \mu) (I_N(u, \mu) - f_j(u, \mu)) \} d\mu du \]

(96) \[= L_T^{(1)} + L_T^{(2)} \quad (say), \]

where
\[ D(u, \mu) = \int_{-\infty}^\infty \{ Q_{j,k}(u, \mu + \frac{x}{M}) - Q_{j,k}(u, \mu) \} W(x) dx. \]

By the dominated convergence theorem,
\[ \lim_{T \to \infty} \| D(u, \mu) \| = 0 \quad a.e. \quad (\mu \in [-\pi, \pi]), \]

therefore, from Lemma 2, \( Var \{ L_T^{(2)} \} = o(1) \), which implies \( L_T^{(2)} = o_p(1) \). On the other hand, by Assumption 3, we have
\[ \int_{-\pi}^\pi f(u, \mu) W_T(\lambda - \mu) d\mu - f(u, \lambda) = O(M^{-2}), \]

hence \( L_T^{(1)} = O \left( \frac{\sqrt{T}}{M^2} \right) = o(1). \)

\( \square \)

**Proof of Theorem 2.** From Assumptions 6, 7 and
\[ f_j(u, \lambda)f_k(u, \lambda)^{-1} = E_m + \frac{(-1)^j}{\sqrt{T}} \Delta(u, \lambda) + O(T^{-1}), \]

it is seen that
\[ D_H(f_j; f_k) = \frac{c}{8\pi T} \int_0^1 \int_{-\pi}^\pi \{ \Delta(u, \lambda) \}^2 d\lambda du + O(T^{-1}) \]

and
\[ V_H^2(j, k) = \frac{c^2}{4\pi T} \int_0^1 \int_{-\pi}^\pi \{ \Delta(u, \lambda) \}^2 d\lambda du + \frac{c^2}{64\pi^4 T} \sum_{a_1, a_2, a_3, a_4=1}^m \kappa_{a_1, a_2, a_3, a_4} \gamma_{a_1, a_2} \gamma_{a_3, a_4} + o(T^{-1}), \]

where \( \gamma_{a,b} \) is the \((a, b)\)-th element of the \( m \times m \) matrix
\[ \Gamma_H = \int_0^1 \int_{-\pi}^\pi A^{(1)}(u, \lambda) \{ A^{(1)}(u, \lambda) \Omega A^{(1)}(u, \lambda)^* \}^{-1} \{ A^{(1)}(u, \lambda) \Omega A^{(2)}(u, \lambda)^* + A^{(2)}(u, \lambda) \Omega A^{(1)}(u, \lambda)^* \} \{ A^{(1)}(u, \lambda) \Omega A^{(1)}(u, \lambda)^* \}^{-1} A^{(1)}(u, \lambda) d\lambda du. \]

If \( \Gamma_H = 0 \), substituting them into (41), then the asymptotic misclassification probabilities are given by
\[ \lim_{T \to \infty} P_D_H(2|1) = \lim_{T \to \infty} P_D_H(1|2) \]

(104) \[= \Phi \left( -\frac{1}{2} \sqrt{\frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \{ \Delta(u, \lambda) \}^2 d\lambda du} \right), \]
Since
\[
\Gamma_H = \int_{-\pi}^{\pi} \{ A^{(2)}(u, \lambda)^* A^{(1)}(u, \lambda)^{-1} \Omega^{-1} + \Omega^{-1} A^{(1)}(u, \lambda)^{-1} A^{(2)}(u, \lambda) \} d\lambda du
\]
\[
(105) = 2\pi \int_{0}^{1} \{ a^{(2)}_0(u) a^{(1)}_0(u)^{-1} \Omega^{-1} + \Omega^{-1} a^{(1)}_0(u)^{-1} a^{(2)}_0(u) \} du,
\]
\[
vec \Gamma_H = 2\pi \int_{0}^{1} vec \{ [K_m + E_m] \Omega^{-1} a^{(1)}_0(u)^{-1} a^{(2)}_0(u) \} du,
\]
which implies that \( \Gamma_H = 0 \) is equivalent to \( a^{(2)}_0(u) \equiv 0 \).
\[
\square
\]

**Proof of Theorem 3.** Under \( \Pi_1 \), it is seen that
\[
\Lambda_T(p_1, p_2) - F_T(p_1, p_2)
\]
\[
= \sum_{t=1}^{T} \left\{ \log p(\varepsilon_t) - \log p \left( \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T} + \sum_{r=0}^{\infty} c^0_{\theta, t, T}(r) \varepsilon - r}{a^0_{\theta, t, T}(0)} \right) \right.
\]
\[
- \log p \left( \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T}}{a^0_{\theta, t, T}(0)} \right) + \log p \left( \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T}}{a^0_{\theta, t, T}(0)} \right) \right\}
\]
\[
= \sum_{t=1}^{T} \left\{ q_{t, T}(\varepsilon_t) + \frac{q^2_{t, T}}{2} D \phi(\varepsilon_t) - r_{t, T} \phi \left( \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T}}{a^0_{\theta, t, T}(0)} \right) \right.
\]
\[\left. - \frac{r^2_{t, T}}{2} D \phi \left( \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T}}{a^0_{\theta, t, T}(0)} \right) \right\} + o_p \left( T^{-1/2} \right)
\]
\[
(107) + O_p \left( T^{-1/2} \right) + o_p \left( 1 \right),
\]
where
\[
q_{t, T} = r_{t, T} + \frac{\sum_{r=0}^{\infty} c^0_{\theta, t, T}(r) \varepsilon - r}{a^0_{\theta, t, T}(0)} - \frac{\sum_{r=0}^{\infty} c^0_{\theta, t, T}(r) \varepsilon - r}{a^0_{\theta, t, T}(0)}
\]
\[\]
\[= r_{t, T} + \frac{h'}{\sqrt{T}} \sum_{r=0}^{\infty} \left\{ \nabla c^0_{\theta, t, T}(r) a^0_{\theta, t, T}(0) + \frac{c^0_{\theta, t, T}(r) \nabla a^0_{\theta, t, T}(0)}{a^0_{\theta, t, T}(0) a^0_{\theta, t, T}(0)} \right\} \varepsilon - r
\]
\[\]
\[= r_{t, T} + O \left( T^{-1/2} t^{-1} \right)
\]
and
\[
r_{t, T} = \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T}}{a^0_{\theta, t, T}(0)} - \frac{\sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) X_{t-k, T}}{a^0_{\theta, t, T}(0)}
\]
\[\]
\[= \frac{h'}{\sqrt{T}} \sum_{k=1}^{t} \left\{ \nabla b^0_{\theta, t, T}(k) a^0_{\theta, t, T}(0) X_{t-k, T} + \sum_{k=0}^{t-1} b^0_{\theta, t, T}(k) \nabla a^0_{\theta, t, T}(0) a^0_{\theta, t, T}(0) X_{t-k, T} \right\}.
\]

\[\]
Here  $\theta^*$, $\theta^{**}$ and $\theta^{***}$ are points on the segment between $\theta$ and $\theta_T = \theta + h/\sqrt{T}$.

From (107), (108) and (109) we can see that $\Lambda_T(p_1, p_2) - F_T(p_1, p_2) = o_P(1)$ under $\Pi_1$. Similarly, we have $\Lambda_T(p_1, p_2) - F_T(p_1, p_2) = o_P(1)$ under $\Pi_2$. Therefore, $F_T(p_1, p_2)$ has the same limit distribution of $\Lambda_T(p_1, p_2)$ under both $\Pi_1$ and $\Pi_2$.

Acknowledgments The author would like to express his sincere thanks to Professor Masanobu Taniguchi for his encouragement and guidance. He is also grateful to a referee for his/her helpful comments and advice.

References


Figure 1: The time varying coefficient functions $b(u)$ (real line) and $b(u)$ (dotted line).

Figure 2: The observation $\{X_{t,T}\}$ of Example 1.
Figure 3: The observation \( \{X_{t,T}\} \) of Example 2.

Figure 4: The time varying spectral density function \( f(u, \lambda) \) for Example 1.
Figure 5: The time varying spectral density function $f(u, \lambda)$ for Example 2.

Figure 6: The observation $\{Y_{t,T}\}$ which is contaminated by a sharp peak.
Figure 7: The time varying spectral density $\tilde{f}_1(u, \lambda)$ which is contaminated by a sharp peak.

Figure 8: The pair of graphs $K(\tilde{f}_1, f_1, f_2)$ (real line) and $K(f_1, f_2)$ (dotted line).
Figure 9: The pair of graphs $B_{1/2}(\bar{f}_1, f_1, f_2)$ (real line) and $B_{1/2}(\bar{f}_1, f_1, f_2)$ (dotted line).